

Guest  
346165.

CAS .

FILE 'REGISTRY' ENTERED AT 09:38:13 ON 26 APR 90  
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT  
COPYRIGHT (C) 1990 AMERICAN CHEMICAL SOCIETY

STRUCTURE FILE UPDATES: HIGHEST RN 126720-44-3  
DICTIONARY FILE UPDATES: 22 APR 90 (900422/ED) HIGHEST RN 126693-39-8

=> d que 12

L2 0 SEA SERYL(2W)PHENYL(2W)ALANYL(2W)CYSTEINYL(2W)ARGINYL(2W)  
PROLYL(2W)ISO(W)LEUCYL

=> d que 13

L3 0 SEA ISO(W)LEUCYL(2W)GLUTAMYL(2W)THREONYL(2W)LEUCYL(2W)VAL  
YL(2W)ASPARTYL(2W)ISO(W)LEUCYL

=> d que 14

L4 0 SEA ALANYL(2W)PROLYL(2W)METHIONYL(2W)ALANYL(2W)GLUTAMYL(2  
W)GLYCYL(2W)GLYCYL

=> d que 15

L5

0 SEA HISTID?(2W)GLUTAMYL(2W)VALYL(2W)VALYL(2W)LYSYL(2W)PHE  
NYL(W)ALANYL(2W)METHIONYL

=>

SUMMARY #Molecular-weight 83478 #Length 737 #Checksum 8125  
SEQUENCE

Initial Score = 6 Optimized Score = 9 Significance = 4.96  
Residue Identity = 35% Matches = 10 Mismatches = 10  
Gaps = 8 Conservative Substitutions = 0

```

      X          10          X
      SFCR--PIEYLVD-----IFQEYYPXXX
          | |  | |  | |  | |  | |
TKRDVNNFDQDFTREEPILTLVDEAIVKQINQEEFKGFSYFGEDLMP
      700          710          720          730
```

10. GUEST-346-1

A22566 3-Phosphoshikimate 1-carboxyvinyltransferase -

ENTRY A22566 #Type Protein  
TITLE 3-Phosphoshikimate 1-carboxyvinyltransferase -  
Salmonella typhimurium #EC-number 2.5.1.19  
ALTERNATE-NAME 5-enolpyruvylshikimate-3-phosphate synthase  
SOURCE Salmonella typhimurium  
ACCESSION A22566  
REFERENCE (Sequence translated from the DNA sequence)  
#Authors Stalker D. M. , Hiatt W. R. , Comai L.  
#Journal J. Biol. Chem. (1985) 260:4724-4728  
#Title A single amino acid substitution in the enzyme  
5-enolpyruvylshikimate-3-phosphate synthase  
confers resistance to the herbicide glyphosate.  
#Comment The authors translated the codon CCT for residue 35  
as Ala.

GENETIC

#Name aroA  
SUMMARY #Molecular-weight 46157 #Length 427 #Checksum 4952  
SEQUENCE

Initial Score = 7 Optimized Score = 9 Significance = 4.96  
Residue Identity = 31% Matches = 10 Mismatches = 10  
Gaps = 12 Conservative Substitutions = 0

```

      X          10          20
      SFCRPIEYLVD-----IFQE-YPXXX
          | |  | |  | |  | |  | |
NEIVLTGEPRMKERPIGHLVDSLRLGGGANIDYLEQENYPPLRLRGGFTGGDI
      120          130          140          150 X          160
```

developmental regulation.  
 COMMENT THIS SEQUENCE HAS NOT BEEN COMPARED TO THE  
 NUCLEOTIDE TRANSLATION.  
 SUMMARY #Molecular-weight 86431 #Length 747 #Checksum 3147  
 SEQUENCE

Initial Score = 6 Optimized Score = 9 Significance = 4.96  
 Residue Identity = 31% Matches = 10 Mismatches = 10  
 Gaps = 12 Conservative Substitutions = 0

```

      X              10              20
      SFCR----PIEY-----LVDIFQEYXXXX
      |||  |||  |||  |||
LCIREKYMQKSFQRFKTPSKYLNRNIDGEALVAIESFYPVFTPPPKKGEDPF
      150  X   160          170          180  X   190
  
```

8. GUEST-346-1  
 A25687

H-2 class II histocompatibility antigen, E-a/k

ENTRY A25687 #Type Protein (fragment)  
 TITLE H-2 class II histocompatibility antigen, E-a/k  
 beta-2 chain precursor - Mouse (fragment)  
 SOURCE Mus musculus #Common-name house mouse  
 ACCESSION A25687  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Braunstein N. S. , Germain R. N.  
 #Journal EMBO J. (1986) 5:2469-2476  
 #Title The mouse E-beta-2 gene: a class II MHC-beta-gene  
 with limited intraspecies polymorphism and an  
 unusual pattern of transcription.

GENETIC  
 #Introns 34/1, 123/1, 217/1

COMMENT THIS SEQUENCE HAS NOT BEEN COMPARED TO THE  
 NUCLEOTIDE TRANSLATION.

SUMMARY #Length 253 #Checksum 8540  
 SEQUENCE

Initial Score = 8 Optimized Score = 9 Significance = 4.96  
 Residue Identity = 30% Matches = 9 Mismatches = 11  
 Gaps = 10 Conservative Substitutions = 0

```

      X              10              20
      SFCRPIEYLVLDIF-----QEYXXXX
      ||  |||  |||  |||
DMLDNYRASVDRCRNNDLVDIFMLNLKAEPKVTVPYPAKTQPLEHHNLLV
      100  X   110          120          130  X   140
  
```

9. GUEST-346-1  
 B28163

Protein kinase C, epsilon type - Rat #EC-number

ENTRY B28163 #Type Protein  
 TITLE Protein kinase C, epsilon type - Rat #EC-number  
 2.7.1-  
 SOURCE Rattus norvegicus #Common-name Norway rat  
 ACCESSION B28163  
 REFERENCE (Sequence translated from the mRNA sequence)  
 #Authors Ono Y. , Fujii T. , Ogita K. , Kikkawa U. , Igarashi K. ,  
 Nishizuka Y.  
 #Journal J. Biol. Chem. (1988) 263:6927-6932  
 #Title The structure, expression, and properties of  
 additional members of the protein kinase C family.  
 SUPERFAMILY #Name protein kinase C  
 KEYWORDS kinase\ phorbol ester receptor\ calcium binding\  
 ATP-binding phosphotransferase

Initial Score = 6 Optimized Score = 9 Significance = 4.96  
Residue Identity = 28% Matches = 10 Mismatches = 10  
Gaps = 15 Conservative Substitutions = 0

```

      X                               10          20
      SFCRPIEY-----LVDIF---QEYPPXXX
      :      :      :      :      :      :
FNSSYRRGDPEFEAMLEYSQGIVDTVAKESLVDIFPWLQIFPNRDLALLKRCLKV
190      200      210      220      230 X      240
  
```

6. GUEST-346-1

GNFFG2 Retrovirus-related pol polyprotein (transposon

ENTRY GNFFG2 #Type Protein (fragment)  
TITLE Retrovirus-related pol polyprotein (transposon  
gypsy) (version 2) - Fruit fly  
DATE 31-Dec-1988 #Sequence 31-Dec-1988 #Text 30-Jun-1989  
PLACEMENT 1451.0 15.0 1.0 1.0 2.0  
SOURCE Drosophila melanogaster  
ACCESSION A23769  
REFERENCE (Sequence translated from the DNA sequence)  
#Authors Yuki S., Ishimaru S., Inouye S., Saigo K.  
#Journal Nucleic Acids Res. (1986) 14:3017-3030  
#Title Identification of genes for reverse  
transcriptase-like enzymes in two Drosophila  
retrotransposons, 412 and gypsy; a rapid detection  
method of reverse transcriptase genes using YXDD  
box probes.  
COMMENT The DNA sequence was obtained from GenBank, release  
54.0.  
SUPERFAMILY #Name pol polyprotein  
KEYWORDS reverse transcriptase\ polyprotein  
SUMMARY #Length 930 #Checksum 7522  
SEQUENCE

Initial Score = 5 Optimized Score = 9 Significance = 4.96  
Residue Identity = 25% Matches = 9 Mismatches = 11  
Gaps = 15 Conservative Substitutions = 0

```

      X                               10          20
      SFCRPIEYL-----VDIFQEYPPXXX
      :      :      :      :      :
NMRVSQEKTRFFKESVEYLGFI VSKDGTKSDPEKVKAIQEYPEPDCVYKVRSLG
350      360      370      380      390 X      400
  
```

7. GUEST-346-1

A27366 AMP deaminase, skeletal muscle - Rat #EC-number

ENTRY A27366 #Type Protein  
TITLE AMP deaminase, skeletal muscle - Rat #EC-number  
3.5.4.6  
ALTERNATE-NAME adenylic acid deaminase\ AMP aminase\ myoadenylate  
deaminase  
SOURCE Rattus norvegicus #Common-name Norway rat  
ACCESSION A27366  
REFERENCE (Sequence translated from the mRNA sequence)  
#Authors Sabina R.L., Marquetant R., Desai N.M., Kaletha K.,  
Holmes E.W.  
#Journal J. Biol. Chem. (1987) 262:12397-12400  
#Title Cloning and sequence of rat myoadenylate deaminase  
cDNA. Evidence for tissue-specific and

BHTLD Hemocyanin d chain - Tarantula (Eurypelma californica)

ENTRY BHTLD #Type Protein

TITLE Hemocyanin d chain - Tarantula (Eurypelma californica)

DATE 15-Nov-1984 #Sequence 15-Nov-1984 #Text 30-Sep-1988

PLACEMENT 646.0 1.0 1.0 1.0 1.0

SOURCE Eurypelma californica

ACCESSION A02565

REFERENCE (Complete sequence)

#Authors Schartau W. , Eyerle F. , Reisinger P. , Geisert H. , Storz H. , Linzen B.

#Journal Hoppe-Seyler's Z. Physiol. Chem. (1983) 364:1383-1409

COMMENT Asn-445 probably binds carbohydrate.

COMMENT Residues 169-177 and 319-327 are thought to form the copper binding site. The two copper ions bound each have 3 nitrogen ligands (presumably contributed by histidine residues) and share a bridging ligand (possibly contributed by a tyrosine residue) in addition to binding oxygen.

COMMENT The hemocyanins are copper-containing, oxygen transport proteins that are highly conserved but found only in arthropods and molluscs. These proteins have a complex and variable quaternary structure with homologous chains aggregating to form either simple hexamers or multihexamer configurations. The tarantula hemocyanin is a 24-chain polymer with seven different chains identified.

SUPERFAMILY #Name hemocyanin

KEYWORDS respiratory protein\ oxygen transport\ copper

SUMMARY #Molecular-weight 72178 #Length 627 #Checksum 9707

SEQUENCE

Initial Score = 6 Optimized Score = 9 Significance = 4.96

Residue Identity = 36% Matches = 9 Mismatches = 11

Gaps = 5 Conservative Substitutions = 0

```

      X           10          20
      SFCRPIEY-----LVDIFQEYPPXXX
      :   : : :   : : : :
NPGVMDDTSTSLRDPIFYRYHRWMDNIFQEYKHLRPSYTHQQLDF
340      350      360      370      X      380

```

#### 5. GUEST-346-1

O4CHC7 Cytochrome P450XVIIA1, steroid 17alpha-monooxygenase

ENTRY O4CHC7 #Type Protein

TITLE Cytochrome P450XVIIA1, steroid 17alpha-monooxygenase - Chicken #EC-number 1.14.99.9

ALTERNATE-NAME cytochrome P450(c17), steroid 17alpha-hydroxylase

DATE 30-Jun-1989 #Sequence 30-Jun-1989 #Text 30-Jun-1989

PLACEMENT 14.0 6.0 1.0 1.0 1.0

SOURCE Gallus gallus #Common-name chicken

ACCESSION JT0318

REFERENCE (Sequence translated from the mRNA sequence)

#Authors Ono H. , Iwasaki M. , Sakamoto N. , Mizuno S.

#Journal Gene (1988) 66:77-85

#Title cDNA cloning and sequence analysis of a chicken gene expressed during the gonadal development and homologous to mammalian cytochrome P-450c17.

#Residues 1-508 (ONO)

SUPERFAMILY #Name cytochrome P450

KEYWORDS steroidogenesis\ ovary\ testis

SUMMARY #Molecular-weight 33663 #Length 353 #Checksum 6372  
SEQUENCE

Initial Score = 9 Optimized Score = 9 Significance = 4.96  
Residue Identity = 40% Matches = 8 Mismatches = 12  
Gaps = 0 Conservative Substitutions = 0

```

      X      10      20
      SFCRPIEYLVDIFQEYPTXXX
      ::  ::  :  :  ::
SVQLRPYN AISFS GPIAVFVS VFLIYPLGQSDWFFPPDFG
  140      X 150      160      X 170
```

### 3. GUEST-346-1

RHRTG Gonadoliberin precursor - Rat

ENTRY RHRTG #Type Protein  
TITLE Gonadoliberin precursor - Rat  
ALTERNATE-NAME gonadotropin releasing hormone\ GnRH\ luteinizing hormone releasing hormone\ LHRH  
INCLUDES gonadotropin releasing hormone\ prolactin release-inhibiting factor  
DATE 31-Mar-1988 #Sequence 31-Mar-1988 #Text 31-Mar-1988  
PLACEMENT 527.0 1.0 2.0 1.0 1.0  
SOURCE Rattus norvegicus #Common-name Norway rat  
ACCESSION B26173  
REFERENCE (Sequence translated from the mRNA sequence)  
#Authors Adelman J. P. , Mason A. J. , Hayflick J. S. , Seeburg P. H.  
#Journal Proc. Nat. Acad. Sci. USA (1986) 83:179-183  
#Title Isolation of the gene and hypothalamic cDNA for the common precursor of gonadotropin-releasing hormone and prolactin release-inhibiting factor in human and rat.  
COMMENT This hormone stimulates the secretion of both luteinizing and follicle-stimulating hormones.  
SUPERFAMILY #Name gonadoliberin  
KEYWORDS reproduction\ prolactin\ amidation\ peptide hormone\ hypothalamus

FEATURE  
1-23 #Domain signal sequence (SIG)\  
24-92 #Protein progonadoliberin (PGN)\  
24-33 #Peptide gonadoliberin (GLN)\  
24 #Modified-site pyrrolidone carboxylic acid, in gonadoliberin (by homology)\  
33 #Modified-site amidated carboxyl end of active gonadoliberin (from Gly-34) (by homology)\  
37-92 #Peptide prolactin release-inhibiting factor (PIF)

SUMMARY #Molecular-weight 10500 #Length 92 #Checksum 1405  
SEQUENCE

Initial Score = 7 Optimized Score = 9 Significance = 4.96  
Residue Identity = 40% Matches = 8 Mismatches = 12  
Gaps = 0 Conservative Substitutions = 0

```

      X      10      20
      SFCRPIEYLVDIFQEYPTXXX
      :  :  ::  ::
SQHWSYGLRPGGKR NTEHLVDSFGEMGKEEDQMAEPQNF
  30      X 40      50      X 60
```

### 4. GUEST-346-1

3.	RAR1G	Guanadine ribonucleoside precursor - Rat	52	7	9	4.96	0
4.	BHTLD	Hemocyanin d chain - Tarantula	627	6	9	4.96	0
5.	Q4CHC7	Cytochrome P450XVIIA1, steroid	508	6	9	4.96	0
6.	GNFFG2	Retrovirus-related pol polypro	930	5	9	4.96	0
7.	A27366	AMP deaminase, skeletal muscle	747	6	9	4.96	0
8.	A25687	H-2 class II histocompatibilit	253	8	9	4.96	0
9.	B28163	Protein kinase C, epsilon type	737	6	9	4.96	0
10.	A22566	3-Phosphoshikimate 1-carboxyvi	427	7	9	4.96	0
**** 3 standard deviations above mean ****							
11.	A24363	Brown fat mitochondrial uncoup	306	8	8	3.72	0
12.	B25687	H-2 class II histocompatibilit	217	8	8	3.72	0
13.	A26294	Uncoupling protein - Rat	307	8	8	3.72	0
14.	HMIVN1	Hemagglutinin precursor - Infl	566	8	8	3.72	0
15.	F2NTD2	Photosystem II D2 protein - Co	353	7	8	3.72	0
16.	F2SPD2	Photosystem II D2 protein - Sp	353	7	8	3.72	0
17.	F2PMD2	Photosystem II D2 protein - Ga	353	7	8	3.72	0
18.	UBBYB	Tubulin beta chain - Yeast (Sa	457	7	8	3.72	0
19.	A27635	Ig heavy chain precursor V reg	122	8	8	3.72	0
20.	A29278	Uncoupling protein - Rat	307	8	8	3.72	0

# 1. GUEST-346-1

A25104 Band 3 protein, nonerythroid (MEB3) - Human

ENTRY A25104 #Type Protein (fragment)  
TITLE Band 3 protein, nonerythroid (MEB3) - Human  
(fragment)  
SOURCE Homo sapiens #Common-name man  
ACCESSION A25104  
REFERENCE (Sequence translated from the mRNA sequence)  
#Authors Demuth D. R. , Showe L. C. , Ballantine M. , Palumbo A. ,  
Fraser P. J. , Cioe L. , Rovera G. , Curtis P. J.  
#Journal EMBO J. (1986) 5:1205-1214  
#Title Cloning and structural characterization of a human  
non-erythroid band 3-like protein.  
SUMMARY #Length 865 #Checksum 7746  
SEQUENCE

Initial Score = 8 Optimized Score = 10 Significance = 6.20  
Residue Identity = 29% Matches = 10 Mismatches = 10  
Gaps = 14 Conservative Substitutions = 0

```

      X                      10                      X
      SFCRPI-----EYLVDIFQEYPXXX
      : : :                      : : : : :
EGSFLVRFVSRFTREIFAFLISLIFYETFYKLVKIFQEHPLHGCSASNSSEVD
    440      X    450          460          470          480

```

# 2. GUEST-346-1

S00929 Photosystem II D2 protein - Barley chloroplast

ENTRY S00929 #Type Protein  
TITLE Photosystem II D2 protein - Barley chloroplast  
SOURCE chloroplast Hordeum vulgare #Common-name barley  
ACCESSION S00929  
REFERENCE (Sequence translated from the DNA sequence)  
#Authors Efimov V. A. , Andreeva A. V. , Reverdatto S. V. ,  
Chakhmakhcheva O. G.  
#Journal Nucleic Acids Res. (1988) 16:5686  
#Title Nucleotide sequence of the barley chloroplast psbD  
gene for the D2 protein of photosystem II.  
#Comment The authors translated the codons GAT for residue  
167 as Gly, CCA for residue 171 as Ala, GAT for  
residue 173 as Ser, and AAA for residue 318 as  
Leu.



## SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.42

```
Times:                CPU                Total Elapsed
                00:01:12.97                00:01:41.00
```

```

Number of residues:          3406022
Number of sequences searched: 12476
Number of scores above cutoff: 4735

```

```
Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
```

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init.	Opt.	Score	Sig.	Frame
**** 4 standard deviations above mean ****							
1. S00929	Photosystem II D2 protein - Ba	353	9	9	4.93	0	
2. HMIVN1	Hemagglutinin precursor - Inf1	566	8	8	4.23	0	
3. CBLV55	Cytochrome b559, component E -	83	8	8	4.23	0	
4. A27817	Lignin peroxidase precursor -	373	8	8	4.23	0	
5. A25539	O-Acetyl homoserine-O-acetyl s	444	8	8	4.23	0	
6. A29278	Uncoupling protein - Rat	307	8	8	4.23	0	
7. A27635	Ig heavy chain precursor V reg	122	8	8	4.23	0	
8. A26294	Uncoupling protein - Rat	307	8	8	4.23	0	
9. A25104	Band 3 protein, nonerythroid (	865	8	10	4.23	0	
10. A24363	Brown fat mitochondrial uncoup	306	8	8	4.23	0	
11. A25687	H-2 class II histocompatibilit	253	8	9	4.23	0	
12. B25687	H-2 class II histocompatibilit	217	8	8	4.23	0	
**** 3 standard deviations above mean ****							
13. F2NTD2	Photosystem II D2 protein - Co	353	7	8	3.52	0	
14. F2LVD2	Photosystem II D2 protein - Li	353	7	8	3.52	0	
15. GVMS11	Ig heavy chain V region - Mous	121	7	7	3.52	0	
16. F2SPD2	Photosystem II D2 protein - Sp	353	7	8	3.52	0	
17. F2PMD2	Photosystem II D2 protein - Ga	353	7	8	3.52	0	
18. UBBYB	Tubulin beta chain - Yeast (Sa	457	7	8	3.52	0	
19. QQBE37	Hypothetical BGLF5 protein - E	470	7	7	3.52	0	
20. HVMS3	Ig heavy chain precursor V reg	117	7	7	3.52	0	

The scores below are sorted by optimized score.  
Significance is calculated based on optimized score.

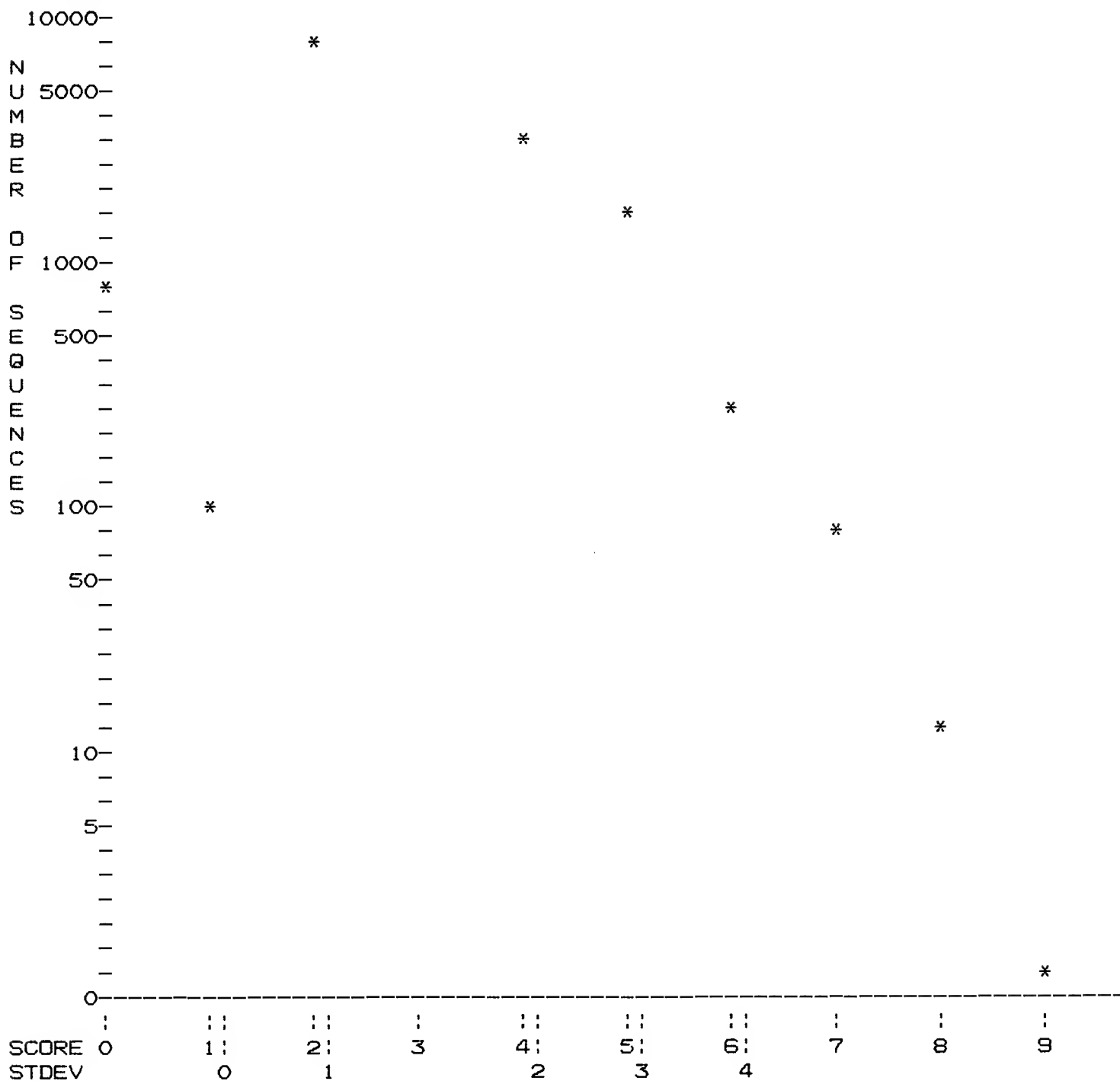
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. A25104	**** 6 standard deviations above mean **** Band 3 protein, nonerythroid (	865	8	10	6.20	0
2. S00929	**** 4 standard deviations above mean **** Photosystem II D2 protein - Ba	353	9	9	4.96	0

Query sequence being compared: GUEST-346-1  
 Number of sequences searched: 12476  
 Number of scores above cutoff: 4735

Results of the initial comparison of GUEST-346-1 with:  
 Data bank : PIR 21.0, all entries



#### PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

Residue Identity = 31% Matches = 10 Mismatches = 10  
 Gaps = 12 Conservative Substitutions = 0

```

      X                      10      20
      SFCR----PIEY-----LVDIFQEYPXXX
      |||  ||  ||  ||  ||  ||  ||  ||
LCIREKYMQKSFQRFPKTPSKYLRNIDGEALVAIESFYPVFTPPPKKGEDPF
    150   X   160           170           180   X   190
  
```

Initial Score = 6    Optimized Score = 9    Significance = 4.74



DR EMBL; M12579; RSGNP1FA.  
 DR PIR; B26173; RHRTG.  
 KW AMIDATION; HORMONE; HYPOTHALAMUS; PLACENTA; SIGNAL.  
 FT SIGNAL 1 23  
 FT CHAIN 24 92 PROGONADOLIBERIN.  
 FT PEPTIDE 24 33 GONADOLIBERIN.  
 FT PEPTIDE 37 92 PROLACTIN RELEASE-INHIBITING FACTOR.  
 FT ACT\_SITE 26 26 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL  
 FT ACTIVITY.  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 92 AA; 10500 MW; 39210 CN;

Initial Score = 7 Optimized Score = 9 Significance = 4.74  
 Residue Identity = 40% Matches = 8 Mismatches = 12  
 Gaps = 0 Conservative Substitutions = 0

X 10 20  
 SFCRPIEYLVDIFQEYPXXX  
 : : : : :  
 SQHWSYGLRPGGKRNTHEHLVDSFQEMGKEEDQMAEPQNF  
 30 X 40 50 X 60

#### 6. GUEST-346-1

AROA\$ECOLI 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2

ID AROA\$ECOLI STANDARD; PRT; 427 AA.  
 AC P07638;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19)  
 DE (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP SYNTHASE)  
 DE (GENE NAME: AROA).  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; BACTERIA; GRAM-NEGATIVE FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA DUNCAN K., LEWENDON A., COGGINS J. R.;  
 RL FEBS LETT. 170:59-63(1984).  
 CC -!- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + 3-PHOSPHOSHIKIMATE =  
 CC ORTHOPHOSPHATE + O(5)-(1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE.  
 CC -!- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS FROM CHORISMATE OF THE  
 CC AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -!- SUBUNIT: MONOMERIC.  
 DR EMBL; X00557; ECAROA.  
 KW AROMATIC AMINO ACID BIOSYNTHESIS; TRANSFERASE.  
 FT ACT\_SITE 408 408 PUTATIVE.  
 SQ SEQUENCE 427 AA; 46164 MW; 892569 CN;

Initial Score = 7 Optimized Score = 9 Significance = 4.74  
 Residue Identity = 31% Matches = 10 Mismatches = 10  
 Gaps = 12 Conservative Substitutions = 0

X 10 20  
 SFCRPIEYLVD-----IFQE-YPXXX  
 : : : : :  
 NDIVLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQGGFTGGNV  
 120 130 140 150 X 160

#### 7. GUEST-346-1

AROA\$SALTY 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2

ID AROA\$SALTY STANDARD; PRT; 427 AA.

SVQLRPYNALISFSGPIAVFVSVFLIYPLGQSDWFFPPDFG  
 140 X 150 160 X 170

#### 4. GUEST-346-1

PSBD\$HORVU PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).

ID PSBD\$HORVU STANDARD; PRT; 353 AA.  
 AC P11849;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
 DE PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).  
 OS BARLEY (HORDEUM VULGARE).  
 OG CHLOROPLAST.  
 OC EUKARYOTA; PLANTA; SPERMATOPHYTA.  
 RN [1] (CV. DONETSKY 6, SEQUENCE FROM N. A.)  
 RA EFIMOV V. A. , ANDREEVA A. V. , REVERDATTO S. V. , JUNG R. ,  
 RA CHAKHMAKHCHEVA O. G. ;  
 RL NUCLEIC ACIDS RES. 16:5686-5686(1988).  
 CC -!- FUNCTION: THIS IS ONE OF THE TWO REACTION CENTRE PROTEINS OF PSII,  
 CC D2 PROTEIN IS NEEDED FOR ASSEMBLY OF A STABLE PSII COMPLEX.  
 CC -!- SIMILARITY: BACTERIAL REACTION CENTER L AND M CHAINS, AND PLANTS  
 CC PHOTOSYSTEM II D1 AND D2 PROTEINS ARE RELATED.  
 DR EMBL; X07522; HVD2PSBD.  
 KW TRANSMEMBRANE; ELECTRON TRANSPORT; THYLAKOID MEMBRANE; PHOTOSYSTEM II;  
 KW CHLOROPLAST; IRON.  
 FT TRANSMEM 36 57  
 FT TRANSMEM 109 129  
 FT TRANSMEM 142 164  
 FT TRANSMEM 192 218  
 FT TRANSMEM 266 286  
 FT METAL 215 215 IRON (NON HAEM).  
 FT METAL 225 225 IRON (NON HAEM).  
 FT METAL 269 269 IRON (NON HAEM).  
 SQ SEQUENCE 353 AA; 39669 MW; 630865 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.74  
 Residue Identity = 40% Matches = 8 Mismatches = 12  
 Gaps = 0 Conservative Substitutions = 0

X 10 20  
 SFCRPIEYLVDIFQEYPXXX  
 :: :: : : ::  
 SVQLRPYNALISFSGPIAVFVSVFLIYPLGQSDWFFPPDFG  
 140 X 150 160 X 170

#### 5. GUEST-346-1

GONL\$RAT GONADOLIBERIN PRECURSOR (LHRH) (LUTEINIZING HORMON

ID GONL\$RAT STANDARD; PRT; 92 AA.  
 AC P07490;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-MAR-1989 (REL. 10, LAST ANNOTATION UPDATE)  
 DE GONADOLIBERIN PRECURSOR (LHRH) (LUTEINIZING HORMONE RELEASING  
 DE HORMONE) (GONADOTROPIN RELEASING HORMONE) (GNRH).  
 OS RAT (RATTUS NORVEGICUS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA ADELMAN J. P. , MASON A. J. , HAYFLICK J. S. , SEEBURG P. H. ;  
 RL PROC. NATL. ACAD. SCI. U. S. A. 83:179-183(1986).  
 CC -!- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND  
 CC FOLLICLE-STIMULATING HORMONES.

KW CHLOROPLAST; IRON.  
 FT TRANSMEM 36 57  
 FT TRANSMEM 109 129  
 FT TRANSMEM 142 164  
 FT TRANSMEM 192 218  
 FT TRANSMEM 266 286  
 FT METAL 215 215 IRON (NON HAEM).  
 FT METAL 225 225 IRON (NON HAEM).  
 FT METAL 269 269 IRON (NON HAEM).  
 SQ SEQUENCE 353 AA; 39571 MW; 631135 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.74  
 Residue Identity = 40% Matches = 8 Mismatches = 12  
 Gaps = 0 Conservative Substitutions = 0

X 10 20  
 SFCRPIEYLVDFQEYPXXX  
 :: :: : : ::  
 SVQLRPYNAISFSGPIAVFVSFLIYPLGQSGWFFAPSFG  
 140 X 150 160 X 170

### 3. GUEST-346-1

PSBD\$ORYSA PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).

ID PSBD\$ORYSA STANDARD; PRT; 353 AA.  
 AC P12095;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
 DE PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).  
 OS RICE (ORYZA SATIVA).  
 OG CHLOROPLAST.  
 OC EUKARYOTA; PLANTA; SPERMATOPHYTA.  
 RN [1] (CV. NIPPONBARE, SEQUENCE FROM N. A.)  
 RA SUGIURA M.;  
 RL SUBMITTED (JUL-1989) TO THE EMBL DATA LIBRARY.  
 RN [2] (GENE ORGANIZATION, SITES, AND FEATURES)  
 RA HIRATSUKA J., SHIMADA H., WHITTIER R., ISHIBASHI T., SAKAMOTO M.,  
 RA MORI M., KONDO C., HONJI Y., SUN C.-R., MENG B.-Y., LI Y.-Q.,  
 RA KANNO A., NISHIZAWA Y., HIRAI A., SHINOZAKI K., SUGIURA M.;  
 RL MOL. GEN. GENET. 217:185-194(1989).  
 CC -!- FUNCTION: THIS IS ONE OF THE TWO REACTION CENTRE PROTEINS OF PSII,  
 CC D2 PROTEIN IS NEEDED FOR ASSEMBLY OF A STABLE PSII COMPLEX.  
 CC -!- SIMILARITY: BACTERIAL REACTION CENTER L AND M CHAINS, AND PLANTS  
 CC PHOTOSYSTEM II D1 AND D2 PROTEINS ARE RELATED.  
 DR EMBL; X15901; CHOSXX.  
 KW TRANSMEMBRANE; ELECTRON TRANSPORT; THYLAKOID MEMBRANE; PHOTOSYSTEM II;  
 KW CHLOROPLAST; IRON.  
 FT TRANSMEM 36 56  
 FT TRANSMEM 109 129  
 FT TRANSMEM 142 164  
 FT TRANSMEM 192 218  
 FT TRANSMEM 266 286  
 FT METAL 215 215 IRON (NON HAEM).  
 FT METAL 225 225 IRON (NON HAEM).  
 FT METAL 269 269 IRON (NON HAEM).  
 SQ SEQUENCE 353 AA; 39573 MW; 631107 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.74  
 Residue Identity = 40% Matches = 8 Mismatches = 12  
 Gaps = 0 Conservative Substitutions = 0

X 10 20  
 SFCRPIEYLVDFQEYPXXX  
 :: :: : : ::



10.	AMDMS\$RAT	AMP DEAMINASE (EC 3.5.4.6) (MY	747	6	9	4.74	0
11.	KPCE\$RAT	PROTEIN KINASE C, EPSILON TYPE	737	6	9	4.74	0
12.	YCY1\$SPIOL	HYPOTHETICAL 250 KD PROTEIN (O	2131	6	9	4.74	0
13.	HCYD\$EURCA	HEMOCYANIN D CHAIN.	627	6	9	4.74	0
**** 3 standard deviations above mean ****							
14.	UCP\$MOUSE	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	3.56	0
15.	LIG1\$PHACH	LIGNINASE PRECURSOR (EC 1.11.1	372	8	8	3.56	0
16.	UCP\$RAT	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	3.56	0
17.	UCP\$MESAU	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	3.56	0
18.	PSBE\$MARPO	CYTOCHROME B559 ALPHA CHAIN (G	83	8	8	3.56	0
19.	MET5\$YEAST	O-ACETYLHOMOSERINE (THIOL)-LYA	444	8	8	3.56	0
20.	HEMA\$INASW	HEMAGGLUTININ PRECURSOR.	566	8	8	3.56	0

# 1. GUEST-346-1

B3LP\$HUMAN NON-ERYTHROID BAND 3-LIKE PROTEIN (HKB3) (FRAGMENT

ID B3LP\$HUMAN STANDARD; PRT; 865 AA.  
AC P04920;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
DT 13-AUG-1987 (REL. 05, LAST ANNOTATION UPDATE)  
DE NON-ERYTHROID BAND 3-LIKE PROTEIN (HKB3) (FRAGMENT).  
OS HUMAN (HOMO SAPIENS).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1] (SEQUENCE FROM N. A.)  
RA DEMUTH D. R. , SHOWE L. C. , BALLANTINE M. , PALUMBO A. , FRASER P. J. ,  
RA CIOE L. , ROVERA G. , CURTIS P. J. ;  
RL EMBO J. 5:1205-1214(1986).  
DR EMBL; X03918; HSHKB3R.  
FT NON\_TER 1 1  
SQ SEQUENCE 865 AA; 95359 MW; 4025588 CN;

Initial Score = 8 Optimized Score = 10 Significance = 5.93  
Residue Identity = 29% Matches = 10 Mismatches = 10  
Gaps = 14 Conservative Substitutions = 0

```

      X                      10                      X
      SFCRPI-----EYLVDIFQEYYPXXX
      : : :                      : : : : :
EGSFLVRFVSRFTREIFAFLISLIFIYETFYKLVKIFQEHPLHGCSASNSSEVD
    440      X    450          460          470          480

```

# 2. GUEST-346-1

PSBD\$SECCE PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).

ID PSBD\$SECCE STANDARD; PRT; 353 AA.  
AC P10803;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
DE PHOTOSYSTEM II D2 PROTEIN (GENE NAME: PSBD).  
OS RYE (SECALE CEREALE).  
OG CHLOROPLAST.  
OC EUKARYOTA; PLANTA; SPERMATOPHYTA.  
RN [1] (SEQUENCE FROM N. A.)  
RA BUKHAROV A. A. , KOLOSOV V. L. , KLEZOVICH O. N. , ZOLOTAREV A. S. ;  
RL NUCLEIC ACIDS RES. 17:798-798(1989).  
CC -!- FUNCTION: THIS IS ONE OF THE TWO REACTION CENTRE PROTEINS OF PSII,  
CC D2 PROTEIN IS NEEDED FOR ASSEMBLY OF A STABLE PSII COMPLEX.  
CC -!- SIMILARITY: BACTERIAL REACTION CENTER L AND M CHAINS, AND PLANTS  
CC PHOTOSYSTEM II D1 AND D2 PROTEINS ARE RELATED.  
DR EMBL; X13366; CHSCPSBD.  
KW TRANSMEMBRANE; ELECTRON TRANSPORT; THYLAKOID MEMBRANE; PHOTOSYSTEM II;

Number of residues: 3757058  
 Number of sequences searched: 12305  
 Number of scores above cutoff: 3754

Cut-off raised to 2.  
 Cut-off raised to 3.  
 Cut-off raised to 4.  
 Cut-off raised to 5.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

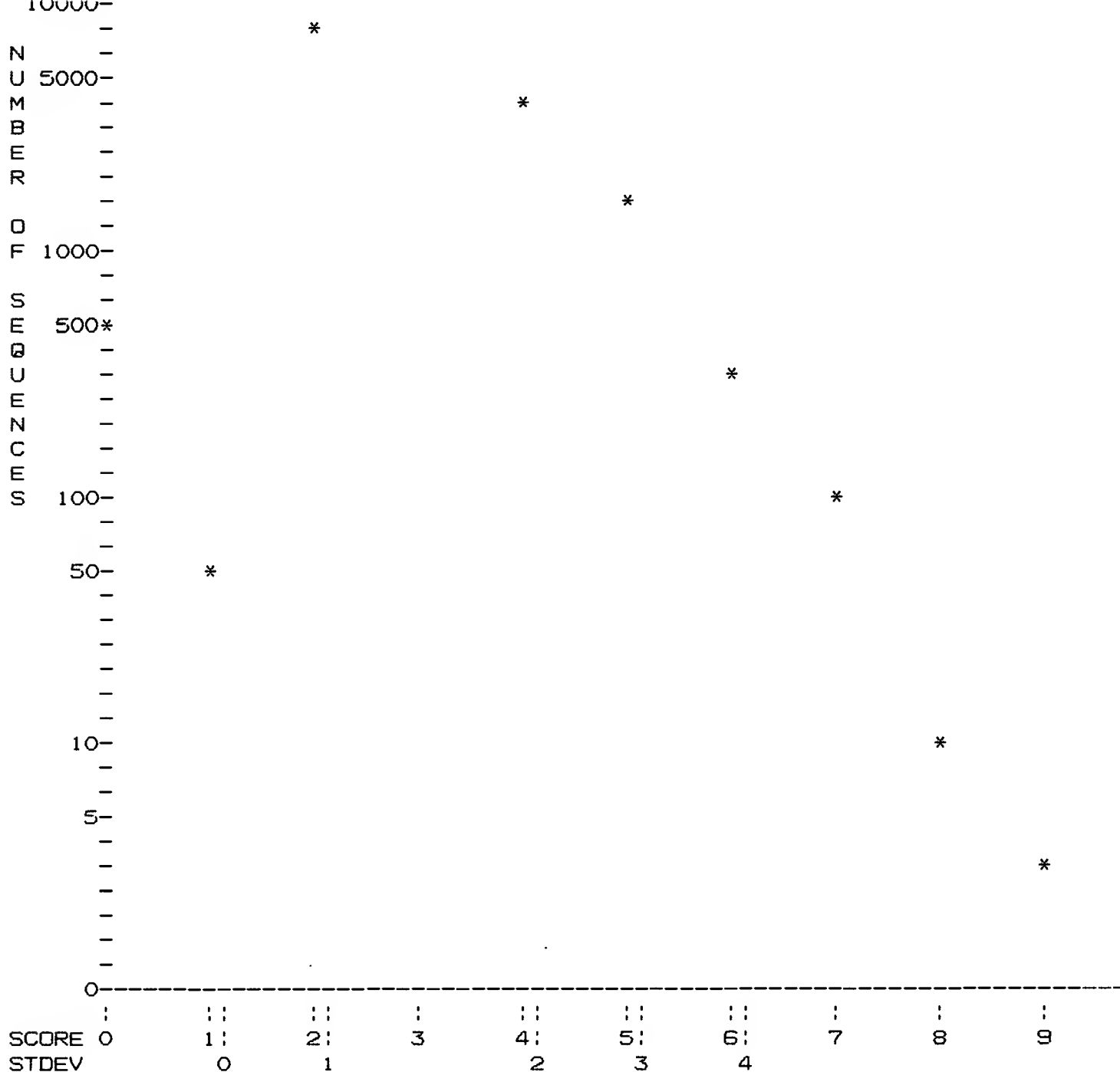
Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. PSBD\$HORVU	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	5.05	0
2. PSBD\$SECCE	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	5.05	0
3. PSBD\$ORYSA	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	5.05	0
**** 4 standard deviations above mean ****						
4. HEMA\$INASW	HEMAGGLUTININ PRECURSOR.	566	8	8	4.33	0
5. LIG1\$PHACH	LIGNINASE PRECURSOR (EC 1.11.1	372	8	8	4.33	0
6. B3LP\$HUMAN	NON-ERYTHROID BAND 3-LIKE PROT	865	8	10	4.33	0
7. MET5\$YEAST	O-ACETYLHOMOSERINE (THIOL)-LYA	444	8	8	4.33	0
8. PSBE\$MARPO	CYTOCHROME B559 ALPHA CHAIN (G	83	8	8	4.33	0
9. UCP\$MESAU	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	4.33	0
10. UCP\$RAT	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	4.33	0
11. UCP\$MOUSE	MITOCHONDRIAL BROWN FAT UNCOUP	306	8	8	4.33	0
**** 3 standard deviations above mean ****						
12. HCYA\$PANIN	HEMOCYANIN A CHAIN.	657	7	7	3.61	0
13. GONL\$RAT	GONADOLIBERIN PRECURSOR (LHRH)	92	7	9	3.61	0
14. ICIC\$HIRME	EGLIN C.	70	7	7	3.61	0
15. HEMA\$INATA	HEMAGGLUTININ (FRAGMENT).	343	7	7	3.61	0
16. HEMA\$INAUS	HEMAGGLUTININ PRECURSOR.	566	7	7	3.61	0
17. HV05\$MOUSE	IG HEAVY CHAIN PRECURSOR V REG	117	7	7	3.61	0
18. HV01\$MOUSE	IG HEAVY CHAIN V REGION (MPC 1	121	7	7	3.61	0
19. HEMA\$INAJP	HEMAGGLUTININ PRECURSOR.	562	7	7	3.61	0
20. CN17\$DICDI	3',5'-CYCLIC-NUCLEOTIDE PHOSPH	452	7	7	3.61	0

The scores below are sorted by optimized score.  
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. B3LP\$HUMAN	NON-ERYTHROID BAND 3-LIKE PROT	865	8	10	5.93	0
**** 4 standard deviations above mean ****						
2. PSBD\$SECCE	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	4.74	0
3. PSBD\$ORYSA	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	4.74	0
4. PSBD\$HORVU	PHOTOSYSTEM II D2 PROTEIN (GEN	353	9	9	4.74	0
5. GONL\$RAT	GONADOLIBERIN PRECURSOR (LHRH)	92	7	9	4.74	0
6. AROA\$ECOLI	3-PHOSPHOSHIKIMATE 1-CARBOXYVI	427	7	9	4.74	0
7. AROA\$SALTY	3-PHOSPHOSHIKIMATE 1-CARBOXYVI	427	7	9	4.74	0
8. CPT1\$CHICK	CYTOCHROME P450 XVIIIA1 (P450-C	508	6	9	4.74	0
9. POL2\$DROME	RETROVIRUS-RELATED POL POLYPRO	930	5	9	4.74	0



#### PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

#### SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.39

Times:	CPU	Total Elapsed
	00:01:51.03	00:01:54.00

Guest

346165

SEQ 1+2  
claim 3

X = any amino acid.

Results file guest-346-1-spt.res made by alexk on Thu 26 Apr 90 11:06:03-PDT.

Query sequence being compared:	GUEST-346-1
Number of sequences searched:	12305
Number of scores above cutoff:	3754

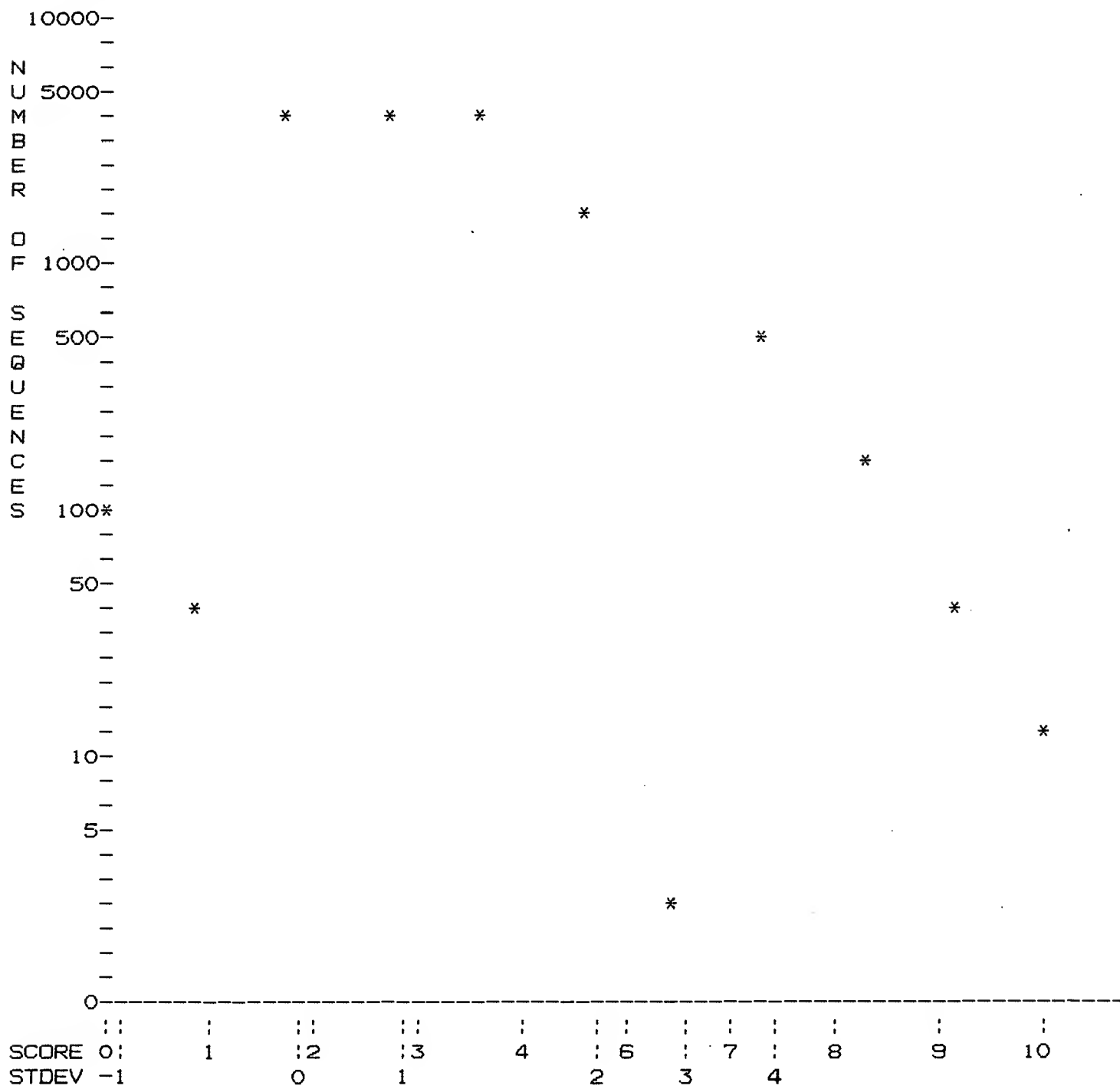
Results of the initial comparison of GUEST-346-1 with:  
Data bank : Swiss-Prot 12, all entries

Guest claim  
346165. 40  
seq 3

Results file guest-346-spt.res made by alexk on Thu 26 Apr 90 9:25:15-PDT.

Query sequence being compared: GUEST-346 -3  
Number of sequences searched: 12305  
Number of scores above cutoff: 3946

Results of the initial comparison of GUEST-346 with:  
Data bank : Swiss-Prot 12, all entries



Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

# SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	4	1.39

Times:	CPU	Total Elapsed
	00:01:16.02	00:05:07.00

Number of residues:	3797058
Number of sequences searched:	12305
Number of scores above cutoff:	3946

Cut-off raised to 2.  
Cut-off raised to 3.  
Cut-off raised to 4.  
Cut-off raised to 5.

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. HIS2\$NEUCR	PHOSPHORIBOSYL-AMP CYCLOHYDROL	863	10	12	5.05	0
2. CFAH\$MOUSE	COMPLEMENT FACTOR H PRECURSOR	1234	10	11	5.05	0
3. IPS\$STRCL	ISOPENICILLIN N SYNTHETASE (IP	329	10	10	5.05	0
4. ODB1\$BOVIN	2-OXOISOVALERATE DEHYDROGENASE	455	10	13	5.05	0
5. ODB1\$HUMAN	2-OXOISOVALERATE DEHYDROGENASE	444	10	13	5.05	0
6. ODB1\$RAT	2-OXOISOVALERATE DEHYDROGENASE	441	10	14	5.05	0
7. PDGA\$HUMAN	PLATELET-DERIVED GROWTH FACTOR	211	10	10	5.05	0
8. PIP1\$BOVIN	1-PHOSPHATIDYLINOSITOL-4,5-BIS	1216	10	12	5.05	0
9. MERA\$STAAU	MERCURIC REDUCTASE (EC 1.16.1.	547	10	11	5.05	0
10. TOXA\$PSEAE	EXOTOXIN A PRECURSOR (EC 2.4.2	638	10	12	5.05	0
11. UL37\$HSV11	PROTEIN UL37 (GENE NAME: UL37)	1123	10	11	5.05	0
**** 4 standard deviations above mean ****						
12. APH6\$ACIBA	APH(3')-VI PROTEIN (3'-AMINOGL	259	9	11	4.33	0
13. CRYT\$BACTI	130 KD CRYSTAL PROTEIN (DELTA	1135	9	11	4.33	0
14. DEDD\$ECOLI	DEDD PROTEIN (GENE NAME: DEDD)	211	9	11	4.33	0
15. FA11\$HUMAN	COAGULATION FACTOR XI PRECURSO	625	9	9	4.33	0
16. LPXA\$ECOLI	UDP-ACETYLGLUCOSAMINE ACYLTRAN	262	9	10	4.33	0
17. CD12\$MOUSE	CD1.2 SURFACE ANTIGEN PRECURSO	297	9	12	4.33	0
18. KS6A\$XENLA	RIBOSOMAL PROTEIN S6 KINASE II	733	9	10	4.33	0
19. KS6B\$XENLA	RIBOSOMAL PROTEIN S6 KINASE II	629	9	10	4.33	0
20. EGF\$HUMAN	EPIDERMAL GROWTH FACTOR (EGF)	1207	9	9	4.33	0

The scores below are sorted by optimized score.  
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. ODP2\$AZOVI	DIHYDROLIPOAMIDE ACETYLTRANSFE	638	9	14	5.87	0
2. ODB1\$RAT	2-OXOISOVALERATE DEHYDROGENASE	441	10	14	5.87	0
3. RADX\$YEAST	DNA REPAIR PROTEIN RAD10 (GENE	210	7	14	5.87	0
**** 4 standard deviations above mean ****						
4. SKI\$HUMAN	SKI ONCOGENE (GENE NAME: SKI).	728	8	13	4.89	0
5. ODB1\$HUMAN	2-OXOISOVALERATE DEHYDROGENASE	444	10	13	4.89	0
6. ODB1\$BOVIN	2-OXOISOVALERATE DEHYDROGENASE	455	10	13	4.89	0
7. DRTS\$LEIMA	DIHYDROFOLATE REDUCTASE (EC 1.	520	4	13	4.89	0
8. H3\$NEUCR	HISTONE H3.	135	7	13	4.89	0
9. PYC\$YEAST	PYRUVATE CARBOXYLASE (EC 6.4.1	1178	8	13	4.89	0
10. VP2\$BT13	VP2 PROTEIN (OUTER CAPSID PROT	959	7	13	4.89	0
**** 3 standard deviations above mean ****						
11. PIP1\$BOVIN	1-PHOSPHATIDYLINOSITOL-4,5-BIS	1216	10	12	3.91	0
12. TOXA\$PSEAE	EXOTOXIN A PRECURSOR (EC 2.4.2	638	10	12	3.91	0
13. KPCE\$RAT	PROTEIN KINASE C, EPSILON TYPE	737	5	12	3.91	0
14. KAD\$MYCCA	ADENYLATE KINASE (EC 2.7.4.3)	213	5	12	3.91	0
15. MOD5\$YEAST	TRNA ISOPENTENYL TRANSFERASE (	427	5	12	3.91	0
16. ODO2\$ECOLI	DIHYDROLIPOAMIDE SUCCINYLTRANS	405	5	12	3.91	0
17. CD12\$MOUSE	CD1.2 SURFACE ANTIGEN PRECURSO	297	9	12	3.91	0
18. HPRT\$SCHMA	HYPOXANTHINE-GUANINE PHOSPHORI	284	5	12	3.91	0
19. CPAX\$HUMAN	CYTOCHROME P450 IIA (EC 1.14.1	489	5	12	3.91	0
20. TRA4\$ECOLI	TRANSPOSASE (TRANSPOSON TN2501	994	5	12	3.91	0

# 1. GUEST-346

ODP2\$AZOVI DIHYDROLIPOAMIDE ACETYLTRANSFERASE COMPONENT (E2)

```

ID  ODP2$AZOVI      STANDARD;      PRT;      638 AA.
AC  P10802;
DT  01-JUL-1989   (REL. 11, CREATED)
DT  01-JUL-1989   (REL. 11, LAST SEQUENCE UPDATE)
DT  01-OCT-1989   (REL. 12, LAST ANNOTATION UPDATE)
DE  DIHYDROLIPOAMIDE ACETYLTRANSFERASE COMPONENT (E2) OF PYRUVATE
DE  DEHYDROGENASE COMPLEX (EC 2.3.1.12).
OS  AZOTOBACTER VINELANDII.
OC  PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI;
OC  AZOTOBACTERIACEAE.
RN  [1] (STRAIN ATCC478, SEQUENCE FROM N. A.)
RA  HANEMAAIJER R., JANSSEN A., DE KOK A., VEEGER C.;
RL  EUR. J. BIOCHEM. 174:593-599(1988).
RN  [2] (LIPOYL DOMAIN CONFORMATION)
RA  HANEMAAIJER R., VERVOORT J., WESTPHAL A.H., DE KOK A., VEEGER C.;
RL  FEBS LETT. 240:205-210(1988).
CC  -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC  CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC  COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC  DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC  (E3).
CC  -!- CATALYTIC ACTIVITY: ACETYL-COA + DIHYDROLIPAMIDE = COA +
CC  S-ACETYLDIHYDROLIPOAMIDE.
CC  -!- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC  SYMMETRY.
CC  -!- COFACTOR: THE E2 COMPONENT CONTAINS THREE COVALENTLY-BOUND LIPOYL
CC  COFACTORS.
CC  -!- THERE ARE THREE COPIES OF THE LIPOYL BINDING DOMAIN.
DR  EMBL; X12455; AVDHAAAT.
KW  GLYCOLYSIS; TRANSFERASE; ACYLTRANSFERASE; DUPLICATION.
FT  DOMAIN      1      327      LIPOYL BINDING.

```





ID RADX\$YEAST STANDARD; PRT; 210 AA.  
 AC P06838;  
 DT 01-JAN-1988 (REL. 06, CREATED)  
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)  
 DE DNA REPAIR PROTEIN RAD10 (GENE NAME: RAD10).  
 OS BAKER'S YEAST (SACCHAROMYCES CEREVISIAE).  
 OC EUKARYOTA; FUNGI; ASCOMYCETES; HEMIASCOMYCETES.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA WEISS W. A. , FRIEDBERG E. C. ;  
 RL EMBO J. 4:1575-1582(1985).  
 RN [2] (CORRECTIONS)  
 RA WEISS W. A. , FRIEDBERG E. C. ;  
 RL EMBO J. 4:3907-3907(1985).  
 RN [3] (SEQUENCE FROM N. A.)  
 RA REYNOLDS P. , PRAKASH L. , DUMAIS D. , PEROZZI G. , PRAKASH S. ;  
 RL EMBO J. 4:3549-3552(1985).  
 CC -!- FUNCTION: THIS PROTEIN IS ONE OF 10 PROTEINS (RAD1, 2,3,4,7,10,14,  
 CC 16,23 & MMS19) INVOLVED IN EXCISION REPAIR OF DNA DAMAGED WITH UV  
 CC LIGHT, BULKY ADDUCTS, OR CROSS-LINKING AGENTS. OF THESE, THE RAD1,  
 CC 2,3,4,10, AND MMS19 PROTEINS SEEM TO BE REQUIRED FOR INCISION OF  
 CC DAMAGED DNA.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SIMILARITY: SOME WITH MAMMALIAN ERCC-1.  
 DR EMBL; X02591; SCRAD10.  
 DR EMBL; X05225; SCRAD10G.  
 KW DNA REPAIR; DNA-BINDING; NUCLEAR PROTEIN.  
 FT DOMAIN 17 23 NUCLEAR LOCATION SIGNAL (PUTATIVE).  
 FT DNA\_BIND 133 153 PUTATIVE.  
 SQ SEQUENCE 210 AA; 24311 MW; 228068 CN;

Initial Score = 7 Optimized Score = 14 Significance = 5.87  
 Residue Identity = 36% Matches = 17 Mismatches = 22  
 Gaps = 8 Conservative Substitutions = 0

```

      X      10      20      30      X
      APMAEGGQKPHE--VVKFMDVYQR----SFXRPIET-LVXIXQ-EYP
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QTSTRRINSNQVINAFNQKPEEWTDSKATDDYNRKRPFRSTRPGKTVLVNTTQKENPLLNLHKSTNW
    50      X 60      70      80      90      100 X 110
  
```

#### 4. GUEST-346

SKI\$HUMAN SKI ONCOGENE (GENE NAME: SKI).

ID SKI\$HUMAN PRELIMINARY; PRT; 728 AA.  
 AC P12755;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
 DE SKI ONCOGENE (GENE NAME: SKI).  
 OS HUMAN (HOMO SAPIENS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA NOMURA N. , SASAMOTO S. , ISHII S. , MATSUI M. , ISHIZAKI R. ;  
 RL NUCLEIC ACIDS RES. 17:5489-5489(1989).  
 CC -!- SIMILARITY: TO SNO ONCOGENE.  
 DR EMBL; X15218; X15218.  
 KW ONCOGENE.  
 SQ SEQUENCE 728 AA; 80004 MW; 2444050 CN;

Initial Score = 8 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 27% Matches = 16 Mismatches = 22  
 Gaps = 20 Conservative Substitutions = 0

```

      X             10             20             30             X
      APMAEGG-----QKPHEVVKFMDVYQ-----RSFXRPIETLVXIXQEYP
            : : : : : : : : : : : : : : : : : : : : : : : :
SGLEAELEHLRQALEGGLDTKEAKEKFLHEVVK-MRVKQEEKLSAALQAKRSLHQELEFLRVAKKEKLREAT
 540      X 550      560      570      580      590      600 X

EAKRNL
610

```

# 5. GUEST-346

ODB1\$HUMAN 2-OXOISOVALERATE DEHYDROGENASE PRECURSOR (EC 1.2.4

```

ID   ODB1$HUMAN      STANDARD;          PRT;    444 AA.
AC   P12694;
DT   01-OCT-1989   (REL. 12, CREATED)
DT   01-OCT-1989   (REL. 12, LAST SEQUENCE UPDATE)
DT   01-OCT-1989   (REL. 12, LAST ANNOTATION UPDATE)
DE   2-OXOISOVALERATE DEHYDROGENASE PRECURSOR (EC 1.2.4.4) (BRANCHED-CHAIN
DE   ALPHA-KETO ACID DEHYDROGENASE COMPONENT (E1)) (FRAGMENT).
OS   HUMAN (HOMO SAPIENS).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1] (SEQUENCE FROM N. A.)
RA   FISHER C. W. , CHUANG J. L. , GRIFFIN T. A. , LAU K. S. , COX R. P. ,
RA   CHUANG D. T. ;
RL   J. BIOL. CHEM. 264:3448-3453(1988).
RN   [2] (LIVER, SEQUENCE OF 67-444 FROM N. A.)
RA   ZHANG B. , CRABB D. W. , HARRIS R. A. ;
RL   GENE 69:159-164(1988).
RN   [3] (MAPLE SYRUP DISEASE MUTATION)
RA   ZHANG B. , EDENBERG H. J. , CRABB D. W. , HARRIS R. A. ;
RL   J. CLIN. INVEST. 83:1425-1429(1989).
CC   -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC   CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
CC   AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS:
CC   BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
CC   ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
CC   -!- CATALYTIC ACTIVITY: 3-METHYL-2-OXOBUTANOATE + LIPOAMIDE =
CC   S-(2-METHYLPROPANOLY)DIHYDROLIPOAMIDE + CO(2).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- DISEASE: MAPLE SYRUP URINE DISEASE IS CAUSED BY A MUTATION AT
CC   POSITION 433.
DR   EMBL; J04474; HSKADA.
DR   EMBL; M22221; HSBCKDH.
KW   OXIDOREDUCTASE; FLAVOPROTEIN; THIAMINE PYROPHOSPHATE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   NON_TER        1          1
FT   TRANSIT        (1        45      MITOCHONDRION.
FT   CHAIN          46        444      2-OXOISOVALERATE DEHYDROGENASE.
FT   VARIANT        433        433      N -> Y (IN MAPLE SYRUP URINE DISEASE).
FT   CONFLICT       247        247      A -> D (IN REF. 2).
SQ   SEQUENCE      444 AA;  50218 MW;  936218 CN;

```

```

Initial Score      =      10   Optimized Score =      13   Significance =   4.89
Residue Identity   =     30%   Matches          =      14   Mismatches   =     25
Gaps               =         7   Conservative Substitutions =         0

```

```

      X             10             20             30             X
      APMAEGGQKPHEVVKFMDVYQ---RSFXRPIETLVXIXQ---EYP
            : : : : : : : : : : : : : : : : : : : : : : : :
KQSRRKVMEAFEQAERKPKPNPNLLFSDVYQEMPAQLRKQGESLARHLQTNGEHYPLDHFQK
 390      X  400      410      420      430      440

```

ID ODB1\$BOVIN STANDARD; PRT; 455 AA.  
 AC P11178;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1989 (REL. 11, LAST ANNOTATION UPDATE)  
 DE 2-OXOISOVALERATE DEHYDROGENASE PRECURSOR (EC 1.2.4.4) (BRANCHED-CHAIN  
 DE ALPHA-KETO ACID DEHYDROGENASE COMPONENT (E1)).  
 OS BOVINE (BOS TAURUS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA HU C.-W. C., LAU K. S., GRIFFIN T. A., CHUANG J. L., FISHER C. W.,  
 RA COX R. P., CHUANG D. T.;  
 RL J. BIOL. CHEM. 263:9007-9014(1988).  
 CC -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX  
 CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA  
 CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS:  
 CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE  
 CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).  
 CC -!- CATALYTIC ACTIVITY: 3-METHYL-2-OXOBUTANEOATE + LIPOAMIDE =  
 CC S-(2-METHYLPROPANOLY)DIHYDROLIPOAMIDE + CO(2).  
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
 DR EMBL; J03759; BTKAD.  
 KW OXIDOREDUCTASE; FLAVOPROTEIN; THIAMINE PYROPHOSPHATE;  
 KW MITOCHONDRION; TRANSIT PEPTIDE.  
 FT TRANSIT 1 55 MITOCHONDRION.  
 FT CHAIN 56 455 2-OXOISOVALERATE DEHYDROGENASE.  
 SQ SEQUENCE 455 AA; 51678 MW; 991502 CN;

Initial Score = 10 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 30% Matches = 13 Mismatches = 26  
 Gaps = 3 Conservative Substitutions = 0

X 10 20 30 X  
 APMAEGGQKPHEVVKFMDVYQ---RSFXRPIETLVXIXQEYP  
 :: :: : :::: : : : :  
 KQSRKKVMEAFEQAERKCLKPNPSLIFSDVYQEMPAQLRKQGESLARHLQTYGEHYPLDHFEEK  
 400 X 410 420 430 440 X 450

7. GUEST-346  
 DRTS\$LEIMA DIHYDROFOLATE REDUCTASE (EC 1.5.1.3) / THYMIDYLATE

ID DRTS\$LEIMA STANDARD; PRT; 520 AA.  
 AC P07382;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-MAR-1989 (REL. 10, LAST ANNOTATION UPDATE)  
 DE DIHYDROFOLATE REDUCTASE (EC 1.5.1.3) / THYMIDYLATE SYNTHASE  
 DE (EC 2.1.1.45).  
 OS LEISHMANIA MAJOR.  
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA BEVERLEY S. M., ELLENBERGER T. E., CORDINGLEY J. S.;  
 RL PROC. NATL. ACAD. SCI. U. S. A. 83:2584-2588(1986).  
 CC -!- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADP(+) =  
 CC 7,8-DIHYDROFOLATE + NADPH.  
 CC -!- CATALYTIC ACTIVITY: 5,10-METHYLENETETRAHYDROFOLATE + DUMP =  
 CC DIHYDROFOLATE + DTMP.  
 CC -!- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,  
 CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DTMP.  
 DR PIR; A23403; RDLNLS.  
 DR EMBL; M12734; LMDHFRTS.

KW MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; TRANSFERASE; NADP;  
 FT DOMAIN 1 233 DIHYDROFOLATE REDUCTASE.  
 FT DOMAIN 234 520 THYMIDYLATE SYNTHASE.  
 FT ACT\_SITE 400 400 BY HOMOLOGY.  
 SQ SEQUENCE 520 AA; 58688 MW; 1352514 CN;

Initial Score = 4 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 29% Matches = 15 Mismatches = 24  
 Gaps = 12 Conservative Substitutions = 0

X 10 20 30 X  
 APMAEG-----GQKPHEVVKFMDVYQRSFXRP-----IET--LVXIXQEYP  
 :: :: : : : : : : : : : :  
 SSKATVEELLAPLPEGQRAAAQDVVVVNGGLAEALRLRLARPLYCSSIETAYCVGGAQVYADAMLSPCIEK  
 110 X 120 130 140 150 160 X 170

# 8. GUEST-346 H3\$NEUCR HISTONE H3.

ID H3\$NEUCR STANDARD; PRT; 135 AA.  
 AC P07041;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-APR-1988 (REL. 07, LAST ANNOTATION UPDATE)  
 DE HISTONE H3.  
 OS NEUROSPORA CRASSA.  
 OC EUKARYOTA; FUNGI; ASCOMYCETES; PYRENO MYCETES.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA WOUTD L. P., PASTINK A., KEMPERS-VEENSTRA A., JANSEN A. E. M.,  
 RA MAGER W. H., PLANTA R. J.;  
 RL NUCLEIC ACIDS RES. 11:5347-5360(1983).  
 DR EMBL; X01612; NCHISH3.  
 KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE.  
 FT INIT\_MET 0 0  
 SQ SEQUENCE 135 AA; 15303 MW; 85124 CN;

Initial Score = 7 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 32% Matches = 17 Mismatches = 21  
 Gaps = 14 Conservative Substitutions = 0

X 10 20 30 X  
 APMAEGGQKPHEVVKFMDV-----YQRS---FXR--PIETLV-XIXQEYP  
 :: : : : : : : : : : : : : : :  
 QLASKAARKSAPSTGGVKKPH-RYKPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKSDLRFQSSAI  
 20 30 40 50 60 70 80

# 9. GUEST-346 PYC\$YEAST PYRUVATE CARBOXYLASE (EC 6.4.1.1) (PYRUVIC CARBOXY

ID PYC\$YEAST STANDARD; PRT; 1178 AA.  
 AC P11154;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
 DE PYRUVATE CARBOXYLASE (EC 6.4.1.1) (PYRUVIC CARBOXYLASE) (PCB) (GENE  
 DE NAME: PYV).  
 OS BAKER'S YEAST (SACCHAROMYCES CEREVISIAE).  
 OC EUKARYOTA; FUNGI; ASCOMYCETES; HEMIASCOMYCETES.  
 RN [1] (SEQUENCE FROM N. A., AND PARTIAL SEQUENCE)  
 RA LIM F., MORRIS C. P., OCCHIODORO F., WALLACE J. C.;  
 RL J. BIOL. CHEM. 263:11493-11497(1988).  
 RN [2] (SEQUENCE OF 1003-1178 FROM N. A.)  
 RA MORRIS C. P., LIM F., WALLACE J. C.;

RL BIOCHEM. BIOPHYS. RES. COMMON. 143:330-336(1987).  
 CC -!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
 CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
 CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE  
 CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND.  
 CC -!- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +  
 CC ORTHOPHOSPHATE + OXALOACETATE.  
 CC -!- PATHWAY: GLUCONEOGENESIS.  
 CC -!- SUBUNIT: TETRAMER.  
 CC -!- COFACTOR: BIOTIN, AND ZINC.  
 CC -!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE  
 CC TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.  
 DR EMBL; J03889; SCPCB.  
 KW LIGASE; MULTIFUNCTIONAL ENZYME; BIOTIN; GLUCONEOGENESIS; ZINC.  
 FT BINDING 1135 1135 BIOTIN (BY SIMILARITY).  
 FT SIMILAR 160 330 CARBAMOYL PHOSPHATE SYNTHETASES.  
 FT SIMILAR 350 470 WITH OTHER BIOTIN CARBOXYLASES.  
 FT SIMILAR 1086 1178 WITH OTHER BIOTIN CARRIER PROTEINS AND  
 FT WITH LIPOAMIDE ACETYLTRANSFERASE.  
 SQ SEQUENCE 1178 AA; 130098 MW; 7059028 CN;

Initial Score = 8 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 31% Matches = 14 Mismatches = 25  
 Gaps = 5 Conservative Substitutions = 0

```

      X          10          20          30          X
      APMAEGGQKPHEVVKFMDV-YQRSFXRPIET----LVXIXQEYP
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    ECDVASYNMYPRVYEDFQKMRETYGDLSVLPTRSFLSPLETDEEIEVVIEQGKTLIIKLQAVGD
    1000  X 1010          1020          1030          1040          1050
  
```

#### 10. GUEST-346

VP2\$BTV13 VP2 PROTEIN (OUTER CAPSID PROTEIN VP2) (GENE NAME:

ID VP2\$BTV13 STANDARD; PRT; 959 AA.  
 AC P12395;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)  
 DE VP2 PROTEIN (OUTER CAPSID PROTEIN VP2) (GENE NAME: L2).  
 OS BLUETONGUE VIRUS (SEROTYPE 13).  
 OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA FUKUSHO A., RITTER G.D., ROY P.;  
 RL J. GEN. VIROL. 68:2967-2973(1987).  
 CC -!- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)  
 CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE  
 CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.  
 DR PIR; A27495; P2XR13.  
 KW COAT PROTEIN.  
 SQ SEQUENCE 959 AA; 112563 MW; 4839211 CN;

Initial Score = 7 Optimized Score = 13 Significance = 4.89  
 Residue Identity = 29% Matches = 14 Mismatches = 25  
 Gaps = 9 Conservative Substitutions = 0

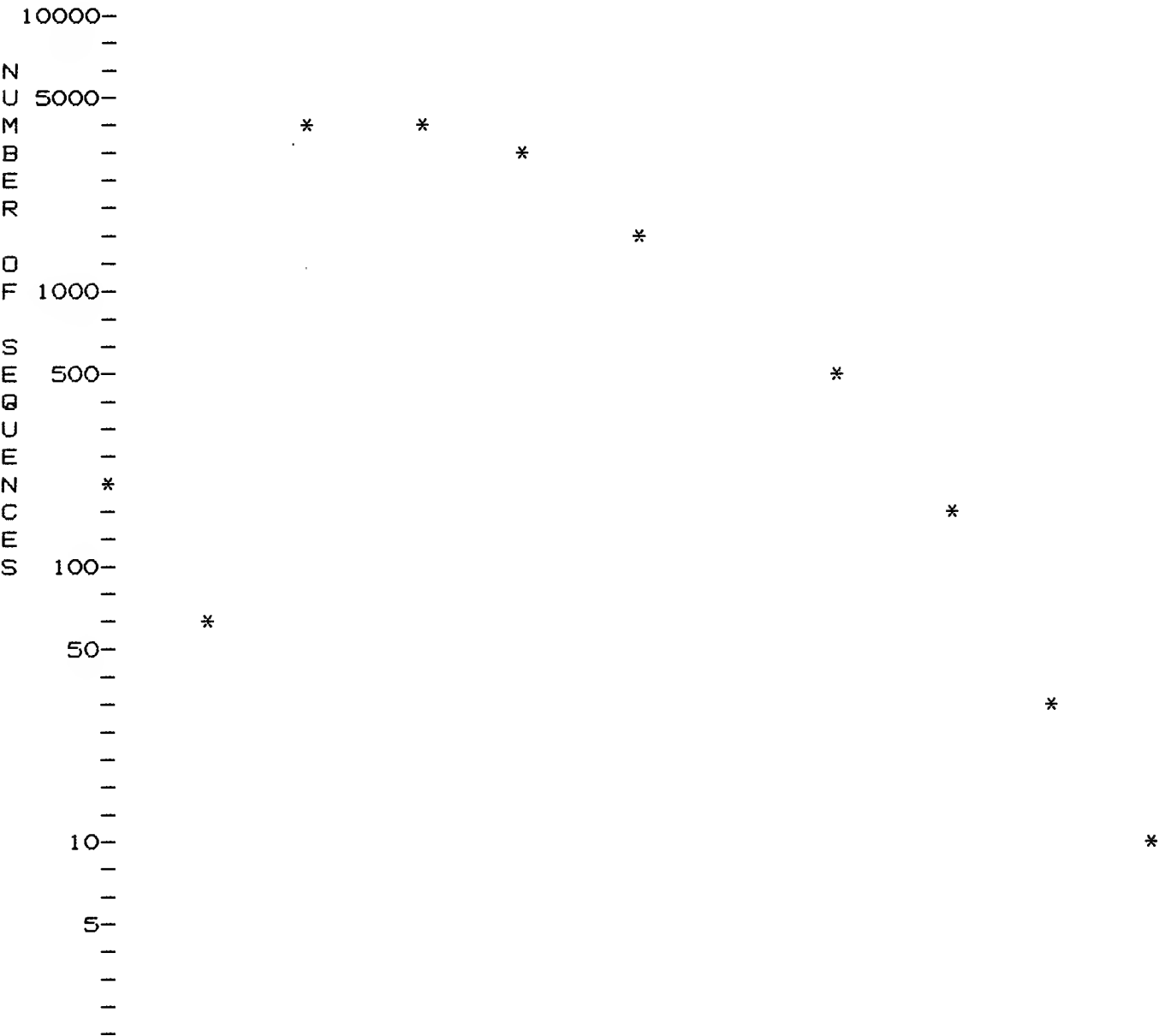
```

      X          10          20          30          X
      APMAEGGQKPHEV----VKFMDVYQRSFXRP-----IETLVXIXQEYP
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    FPPYFDQWTVVPMFNARIKPCEVEVGERKNIDPYVKRTHRPLKADCIELMRYHMSQYMDLRVSLQGT
    520  X 530          540          550          560          570
  
```

Results file guest-346.res made by alexk on Thu 26 Apr 90 9:20:03-PDT.

Query sequence being compared:	GUEST-346
Number of sequences searched:	12476
Number of scores above cutoff:	3893

Results of the initial comparison of GUEST-346 with:  
Data bank : PIR 21.0, all entries



```

-
0-----*-----
SCORE 0: 1 2 3 4 6 7 8 9 10
STDEV -1 0 1 2 3 4

```

# PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      1      Joining penalty      20
Gap penalty      1.00      Window size      32
Gap size penalty      0.05
Cutoff score      0
Randomization group      0

Initial scores to save      20      Alignments to save      10
Optimized scores to save      20      Display context      10

```

# SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              3          4          1.39

Times:      CPU      Total Elapsed
            00:01:12.02      00:04:55.00

Number of residues:      3406022
Number of sequences searched:      12476
Number of scores above cutoff:      3893

```

Cut-off raised to 2.  
 Cut-off raised to 3.  
 Cut-off raised to 4.  
 Cut-off raised to 5.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. SHNC	Phosphoribosyl-AMP cyclohydrol	863	10	12	5.03	0
2. NBMSH	Complement factor H precursor	1234	10	11	5.03	0
3. PFHUG1	Platelet-derived growth factor	211	10	10	5.03	0
4. E29504	Mercuric reductase - Staphyloc	547	10	11	5.03	0
5. B28964	Platelet-derived growth factor	196	10	10	5.03	0
6. A28073	2-Oxoisovalerate dehydrogenase	455	10	13	5.03	0
7. A28964	Platelet-derived growth factor	211	10	10	5.03	0
8. A29468	2-Oxoisovalerate dehydrogenase	441	10	14	5.03	0
**** 4 standard deviations above mean ****						
9. XUECDP	UDP-acetylglucosamine acyltran	115	9	10	4.32	0
10. SHBY	Phosphoribosyl-AMP cyclohydrol	799	9	10	4.32	0
11. DTECC	Aspartate carbamoyltransferase	311	9	10	4.32	0
12. KFHU1	Coagulation factor XI precurs	625	9	9	4.32	0
13. OGHU	Hemopexin precursor - Human (f	441	9	10	4.32	0
14. XMECDD	dedD protein - Escherichia col	211	9	11	4.32	0
15. W2WLDP	Probable E2 protein - Deer pap	416	9	10	4.32	0
16. P5XR10	Outer capsid protein VP5 - Blu	526	9	9	4.32	0

17. Z2BPC2	Gene 12 protein - bacteriophag	458	9	10	4.32	0
18. A23162	Extension - Carrot (fragment)	154	9	10	4.32	0
19. JT0315	Parasporal crystal protein - B	1135	9	11	4.32	0
20. S00049	Aspartate carbamoyltransferase	311	9	10	4.32	0

The scores below are sorted by optimized score.  
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. A29468	2-Oxoisovalerate dehydrogenase	441	10	14	5.96	0
2. XYECCR	Chemotaxis protein methylase -	286	4	14	5.96	0
3. B22726	RAD10 protein - Yeast (Sacchar	195	7	14	5.96	0
4. A24576	RAD10 protein - Yeast (Sacchar	210	7	14	5.96	0
**** 4 standard deviations above mean ****						
5. VVVPC2	Coat proteins VP2 and VP3 - Mo	319	4	13	4.97	0
6. A28073	2-Oxoisovalerate dehydrogenase	455	10	13	4.97	0
7. VVVP2	Coat proteins VP2 and VP3 - Mo	319	4	13	4.97	0
8. RDLNTS	Dihydrofolate reductase/thymid	520	4	13	4.97	0
9. A29233	Pyruvate carboxylase - Yeast (	1178	8	13	4.97	0
10. A05032	Hypothetical protein 548 (homo	548	7	13	4.97	0
11. P2XR13	VP2 protein - Bluetongue virus	959	7	13	4.97	0
12. B28814	Ig heavy chain V region - Chic	116	7	13	4.97	0
13. A28912	Kinase-related protein sevenle	2554	5	13	4.97	0
**** 3 standard deviations above mean ****						
14. S00373	Histone H3 - Wheat	135	7	12	3.97	0
15. GNWVY	Genome polyprotein - Yellow fe	3411	7	12	3.97	0
16. A25564	Histone H3 - Rice	136	7	12	3.97	0
17. A26014	Histone H3 - Wheat	136	7	12	3.97	0
18. HVMS3	Ig heavy chain precursor V reg	117	7	12	3.97	0
19. A05129	Cholera enterotoxin, A chain p	258	7	12	3.97	0
20. A27126	Multidrug resistance protein 1	572	7	12	3.97	0

1. GUEST-346  
A29468 2-Oxoisovalerate dehydrogenase (lipoamide), E1- $\alpha$ p

ENTRY A29468 #Type Protein (fragment)  
TITLE 2-Oxoisovalerate dehydrogenase (lipoamide), E1- $\alpha$  chain precursor - Rat (fragment) #EC-number 1.2.4.4  
ALTERNATE-NAME branched-chain  $\alpha$ -keto acid dehydrogenase  
SOURCE Rattus norvegicus #Common-name Norway rat  
ACCESSION A29468  
REFERENCE (Sequence translated from the mRNA sequence)  
#Authors Zhang B. , Kuntz M. J. , Goodwin G. W. , Harris R. A. , Crabb D. W.  
#Journal J. Biol. Chem. (1987) 262:15220-15224  
#Title Molecular cloning of a cDNA for the E1- $\alpha$  subunit of rat liver branched chain  $\alpha$ -ketoacid dehydrogenase.

SUMMARY #Length 441 #Checksum 8882  
SEQUENCE

Initial Score = 10 Optimized Score = 14 Significance = 5.96  
Residue Identity = 33% Matches = 14 Mismatches = 25  
Gaps = 3 Conservative Substitutions = 0

X 10 20 30 X



KQSRKKVMEAFEQAERKLLKPNPSSLFSDVYQEMPAQLRRQQESLARHLQTYGEHYPLDHFQK  
 380 390 400 410 420 430 440

## 2. GUEST-346

XYECCR Chemotaxis protein methylase - Escherichia coli

ENTRY XYECCR #Type Protein  
 TITLE Chemotaxis protein methylase - Escherichia coli  
 #EC-number 2.1.1.-  
 DATE 28-Dec-1987 #Sequence 28-Dec-1987 #Text 28-Dec-1987  
 PLACEMENT 127.0 1.0 1.0 1.0 1.0  
 SOURCE Escherichia coli  
 ACCESSION C25195  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Mutoh N., Simon M. I.  
 #Journal J. Bacteriol. (1986) 165:161-166  
 #Title Nucleotide sequence corresponding to five chemotaxis  
 genes in Escherichia coli.  
 COMMENT This protein appears to be a methylesterase  
 specifically responsible for removing the methyl  
 group from the gamma-glutamyl methyl ester  
 residues in the methyl-accepting chemotaxis  
 proteins (MCP). The MCP methylation state of the  
 cell is crucial for sensory responses and  
 adaptations.  
 GENETIC  
 #Map-position 42  
 #Name cheR  
 SUPERFAMILY #Name chemotaxis protein methylase  
 KEYWORDS methyltransferase\ chemotaxis response  
 SUMMARY #Molecular-weight 32750 #Length 286 #Checksum 10  
 SEQUENCE

Initial Score = 4 Optimized Score = 14 Significance = 5.96  
 Residue Identity = 34% Matches = 15 Mismatches = 24  
 Gaps = 5 Conservative Substitutions = 0

X 10 20 30 X  
 APMAEGGQKPEHVVK--FMDVYQRSFXR---PIETLVXIXQEYF  
 : : : : : : : : : : : : : : : :  
 VFASDIDTEVLEKARSGIYRHEELKNLTPQQLQRYFMRGTGPHEGLVRVRQELANYVDFAPLNL  
 150 160 170 180 190 200 X 210

## 3. GUEST-346

B22726 RAD10 protein - Yeast (Saccharomyces cerevisiae)

ENTRY B22726 #Type Protein  
 TITLE RAD10 protein - Yeast (Saccharomyces cerevisiae)  
 SOURCE Saccharomyces cerevisiae  
 ACCESSION B22726  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Weiss W. A., Friedberg E. C.  
 #Journal EMBO J. (1985) 4:1575-1582  
 #Title Molecular cloning and characterization of the yeast  
 RAD10 gene and expression of RAD10 protein in E.  
 coli.  
 GENETIC  
 #Map-position 8R  
 #Name RAD10  
 SUMMARY #Molecular-weight 22614 #Length 195 #Checksum 1203  
 SEQUENCE

Initial Score = 7 Optimized Score = 14 Significance = 5.96  
 Residue Identity = 36% Matches = 17 Mismatches = 22  
 Gaps = 8 Conservative Substitutions = 0

```

      X      10      20      30      X
      APMAEGGQKPHE--VVKFMDVYQR----SFXRPIET-LVXIXQ-EYP
            :   : : : :   :   : : :   :   : : :   :   :
QTSTRRINSNQVINAFNQKPEEWTDKATDDYNRKRPFRSTRPGKTVLVNTTQKENPLLNLHLKSTNW
      50      X 60      70      80      90      100 X 110
  
```

#### 4. GUEST-346

A24576 RAD10 protein - Yeast (*Saccharomyces cerevisiae*)

ENTRY A24576 #Type Protein  
 TITLE RAD10 protein - Yeast (*Saccharomyces cerevisiae*)  
 SOURCE *Saccharomyces cerevisiae*  
 ACCESSION A24576  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Reynolds P. , Prakash L. , Dumais D. , Perozzi G. ,  
 Prakash S.  
 #Journal EMBO J. (1985) 4:3549-3552  
 GENETIC  
 #Name RAD10  
 SUPERFAMILY #Name Gene RAD10 protein  
 KEYWORDS UV  
 SUMMARY #Molecular-weight 24311 #Length 210 #Checksum 5515  
 SEQUENCE

Initial Score = 7 Optimized Score = 14 Significance = 5.96  
 Residue Identity = 36% Matches = 17 Mismatches = 22  
 Gaps = 8 Conservative Substitutions = 0

```

      X      10      20      30      X
      APMAEGGQKPHE--VVKFMDVYQR----SFXRPIET-LVXIXQ-EYP
            :   : : : :   :   : : :   :   : : :   :   :
QTSTRRINSNQVINAFNQKPEEWTDKATDDYNRKRPFRSTRPGKTVLVNTTQKENPLLNLHLKSTNW
      50      X 60      70      80      90      100 X 110
  
```

#### 5. GUEST-346

VVVP2 Coat proteins VP2 and VP3 - Mouse polyomavirus

ENTRY VVVP2 #Type Protein  
 TITLE Coat proteins VP2 and VP3 - Mouse polyomavirus  
 (strain Crawford small-plaque)  
 DATE 30-Jun-1989 #Sequence 30-Jun-1989 #Text 30-Jun-1989  
 PLACEMENT 1163.0 3.0 1.0 1.0 2.0  
 SOURCE mouse polyomavirus  
 ACCESSION E28838  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Rothwell V.M. , Folk W.R.  
 #Journal J. Virol. (1983) 48:472-480  
 #Title Comparison of the DNA sequence of the Crawford  
 small-plaque variant of polyomavirus with those of  
 polyomaviruses A2 and strain 3.  
 #Residues 1-319 (R01)  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Rothwell V.M.  
 #Citation submitted to GenBank, November 1985  
 #Residues 1-319 (R02)  
 COMMENT The DNA sequence was obtained from GenBank, release  
 56.0.  
 COMMENT This virus is a member of the family Papovaviridae.  
 SUPERFAMILY #Name polyoma coat proteins VP2 and VP3  
 KEYWORDS late protein

```

FEATURE
1-319          #Protein coat protein VP2 (VP2)\
116-319        #Protein coat protein VP3 (VP3)
SUMMARY        #Molecular-weight 34827 #Length 319 #Checksum 2781
SEQUENCE

```

```

Initial Score      =      4   Optimized Score =      13   Significance =   4.97
Residue Identity   =     28%   Matches          =      14   Mismatches   =     25
Gaps               =      10   Conservative Substitutions =      0

```

```

      X      10      20      30      X
      APMAEGGQKP-----HEVVKFMDVYGRSFXR--PIETLVXIXQEYP
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QRRALFNRIEGSMGNGGPTPAAHIQDESGEVIKFYQAGVVSHQRVTPDWMLPLILGLYGDITPTWATVI
240      250      260      270      280      290      300

```

6. GUEST-346  
A28073      2-Oxoisovalerate dehydrogenase (lipoamide), E1 alp

```

ENTRY          A28073      #Type Protein
TITLE          2-Oxoisovalerate dehydrogenase (lipoamide), E1 alpha
               chain precursor - Bovine #EC-number 1.2.4.4
ALTERNATE-NAME branched-chain alpha-keto acid dehydrogenase E1
               alpha chain\ branched-chain alpha-keto acid
               decarboxylase
SOURCE         Bos primigenius taurus #Common-name cattle
ACCESSION      A28073
REFERENCE      (Sequence translated from the mRNA sequence)
               #Authors    Hu C.W.C. , Lau K.S. , Griffin T.A. , Chuang J.L. ,
               Fisher C.W. , Cox R.P. , Chuang D.T.
               #Journal    J. Biol. Chem. (1988) 263:9007-9014
               #Title      Isolation and sequencing of a cDNA encoding the
               decarboxylase (E1)-alpha precursor of bovine
               branched-chain alpha-keto acid dehydrogenase
               complex. Expression of E1-alpha mRNA and subunit
               in maple-syrup-urine-disease and 3T3-L1 cells.

```

```

FEATURE
56-455          #Protein 2-oxoisovalerate dehydrogenase
               (lipoamide), E1 alpha chain (MAT)
SUMMARY        #Molecular-weight 51678 #Length 455 #Checksum 4630
SEQUENCE

```

```

Initial Score      =     10   Optimized Score =     13   Significance =   4.97
Residue Identity   =     30%   Matches          =     13   Mismatches   =     26
Gaps               =       3   Conservative Substitutions =     0

```

```

      X      10      20      30      X
      APMAEGGQKPHEVVKFMDVYQ---RSFXRPIETLVXIXQEYP
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
KQSRKKVMEAFEQAERKCLKPNPSLIFSVDVYQEMPAQLRKQGESLARHLQTYGEHYPLDHFEEK
400      X  410      420      430      440      X  450

```

7. GUEST-346  
VVVP2      Coat proteins VP2 and VP3 - Mouse polyomavirus

```

ENTRY          VVVP2      #Type Protein
TITLE          Coat proteins VP2 and VP3 - Mouse polyomavirus
DATE          31-Jul-1980 #Sequence 08-Oct-1981 #Text 27-Nov-1985
PLACEMENT     1163.0      3.0      1.0      1.0      1.0
SOURCE         mouse polyomavirus
ACCESSION      A03635
COMMENT        The VP2 sequence of strain A2 is shown; VP3
               corresponds to residues 116-319.
REFERENCE      (Strain A2, sequence translated from the DNA

```

```

sequence)
#Authors      Soeda E., Accard J. R., Smolar N., Walsh J. E.,
               Griffin B. E.
#Journal      Nature (1980) 283:445-453
REFERENCE     (Strain 3, sequence translated from the DNA
               sequence)
#Authors      Friedmann T., Esty A., LaPorte P., Deininger P.
#Journal      Cell (1979) 17:715-724
#Comment      This sequence differs from that shown in having
               78-Asn, 219-Val, 276-Pro, 277-Gly, 278-Gly, and
               279-Ala.
SUMMARY       #Molecular-weight 34800 #Length 319 #Checksum 2886
SEQUENCE

```

```

Initial Score      =      4   Optimized Score =      13   Significance =   4.97
Residue Identity   =     28%   Matches          =      14   Mismatches   =     25
Gaps               =      10   Conservative Substitutions =      0

```

```

      X          10          20          30          X
      APMAEGGQKP-----HEVVKFMDVYQRSFXR--PIETLVXIXQEYP
          |  |  |          |  |  |          |  |  |          |
QRRALFNRIEGSMGNGGPTPAAHIQDESGEVIKFYQAGVVSHQRVTPDWMLPLILGLYGDITPTWATVI
240      250      260      270      280      290      300

```

# 8. GUEST-346 RDLNTS      Dihydrofolate reductase/thymidylate synthase -

```

ENTRY          RDLNTS      #Type Protein
TITLE          Dihydrofolate reductase/thymidylate synthase -
               Leishmania tropica #EC-number 1.5.1.3 #EC-number
               2.1.1.45
DATE          28-Dec-1987 #Sequence 28-Dec-1987 #Text 31-Mar-1988
PLACEMENT     128.0      2.0      1.0      1.0      1.0
SOURCE        Leishmania tropica major
ACCESSION     A23403
REFERENCE     (Sequence translated from the DNA sequence)
#Authors      Beverley S.M., Ellenberger T.E., Cordingley J.S.
#Journal      Proc. Nat. Acad. Sci. USA (1986) 83:2584-2588
#Title        Primary structure of the gene encoding the
               bifunctional dihydrofolate reductase-thymidylate
               synthase of Leishmania major.
SUPERFAMILY    1-520 #Name DHFR-TS bifunctional enzyme\
               1-233 #Name dihydrofolate reductase\
               234-520 #Name thymidylate synthase
KEYWORDS      bifunctional enzyme\ oxidoreductase\ synthase
SUMMARY       #Molecular-weight 58688 #Length 520 #Checksum 2419
SEQUENCE

```

```

Initial Score      =      4   Optimized Score =      13   Significance =   4.97
Residue Identity   =     29%   Matches          =      15   Mismatches   =     24
Gaps               =      12   Conservative Substitutions =      0

```

```

      X          10          20          30          X
      APMAEG-----GQKPHEVVKFMDVYQRSFXRP-----IET--LVXIXQEYP
          |  |  |          |  |  |          |  |  |          |
SSKATVEELLAPLPEGQRAAAAQDVVVVNGGLAEALRLLARPLYCSSIETAYCVGGAGVYADAMLSPCIEK
110 X    120      130      140      150      160 X    170

```

# 9. GUEST-346 A29233      Pyruvate carboxylase - Yeast (Saccharomyces

```

ENTRY          A29233      #Type Protein
TITLE          Pyruvate carboxylase - Yeast (Saccharomyces
               cerevisiae) #EC-number 6.4.1.1

```

ALTERNATE-NAME pyruvic carboxylase  
 SOURCE Saccharomyces cerevisiae  
 ACCESSION A29233\ A29722  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Lim F. , Morris C. P. , Occhiodoro F. , Wallace J. C.  
 #Journal J. Biol. Chem. (1988) 263:11493-11497  
 #Title Sequence and domain structure of yeast pyruvate  
 carboxylase.  
 SUMMARY #Molecular-weight 130098 #Length 1178 #Checksum 6631  
 SEQUENCE

Initial Score = 8 Optimized Score = 13 Significance = 4.97  
 Residue Identity = 31% Matches = 14 Mismatches = 25  
 Gaps = 5 Conservative Substitutions = 0

```

      X      10      20      30      X
      APMAEGGQKPHEVVKFMDV-YQRSFXRPIET----LVXIXQEYP
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    ECDVASYNMYPRVYEDFQKMRETYGDLSVLPTRSFLSPLETDEEIEVVIEQGKTLIIKLQAVGD
    1000  X  1010      1020      1030      1040      1050
  
```

# 10. GUEST-346

A05032 Hypothetical protein 548 (homolog of E. coli rpoC)

ENTRY A05032 #Type Protein  
 TITLE Hypothetical protein 548 (homolog of E. coli rpoC) -  
 Common tobacco chloroplast  
 SOURCE chloroplast Nicotiana tabacum #Common-name common  
 tobacco  
 ACCESSION A05032  
 REFERENCE (cv. Bright Yellow 4, sequence translated from the  
 DNA sequence)  
 #Authors Sugiura M.  
 #Citation submitted to EMBL, August 1986, in computer-readable  
 form  
 REFERENCE (cv. Bright Yellow 4; gene organization, sites, and  
 features)  
 #Authors Shinozaki K. , Ohme M. , Tanaka M. , Wakasugi T. ,  
 Hayashida N. , Matsubayashi T. , Zaita N. ,  
 Chunwongse J. , Obokata J. , Yamaguchi-Shinozaki K. ,  
 Ohto C. , Torazawa K. , Meng B. Y. , Sugita M. , Deno  
 H. , Kamogashira T. , Yamada K. , Kusuda J. , Takaiwa  
 F. , Kato A. , Tohdoh N. , Shimada H. , Sugiura M.  
 #Journal EMBO J. (1986) 5:2043-2049  
 GENETIC  
 #Start-codon AGG  
 COMMENT The code is Q5NT48.  
 SUMMARY #Molecular-weight 63034 #Length 548 #Checksum 2349  
 SEQUENCE

Initial Score = 7 Optimized Score = 13 Significance = 4.97  
 Residue Identity = 35% Matches = 14 Mismatches = 22  
 Gaps = 4 Conservative Substitutions = 0

```

      X      10      20      30      X
      APMAEGGQKPHEVVK-FMDVYQRSFXRPIETLVXIXQEYP
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    DTLLDNGIRGQPMRDGHNK---VYKSFSVDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPS
    200      210      220      230      240      X 250
  
```